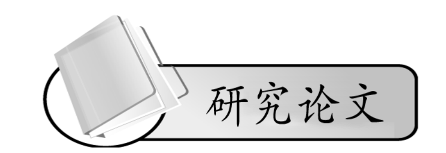


基于被引科学知识主题突变的突破性创新识别*

张金柱[1] 张晓林[2]
[1](南京理工大学经济管理学院 南京 210094)
[2](中国科学院文献情报中心 北京 100190)

摘要:【目的】在与基础研究密切相关的技术领域, 需要从专利信息中的被引科学知识主题内容出发识别突破性创新。【方法】抽取专利科学引文的关键词和学科分类表示被引科学知识; 在关键词共现网络和学科分类组合中识别被引科学知识的主题; 提出基于关键词和学科分类的主题突变程度计算方法, 遴选突变程度高的主题对应作为突破性创新发生的技术主题, 对突破性创新进行识别。【结果】在纳米电子学领域识别出已被证实为突破性创新的纳米电路相关主题, 即纳米导线、碳纳米管、可计算电路等纳米电路材料和制备主题, 并且是材料科学、化学、光学、生物学和应用物理等多学科交叉融会的结果, 验证了该方法的有效性。【局限】被引科学知识抽取、预处理和匹配准确率需要提高, 方法通用性还需在其他领域进行验证。【结论】该方法是基于专利信息识别突破性创新的重要完善和补充, 可扩展应用到其他与基础研究密切相关的技术领域中识别突破性创新。

关键词: 突破性创新 被引科学知识 主题突变 突变程度 纳米电子学
分类号: G305 G353

1 引 言

突破性创新(Radical Innovation, RI)是指技术创新的方法、产品、设备、材料等技术主题发生不连续性变化, 并引发性能的跃迁或功能的变化, 最终导致市场、产品、服务、商业模式等发生不连续性变化[1-2]。揭示和判别可能发生突破性创新的领域和主题, 对规划技术发展方向和优先主题、规避潜在落后技术、优化研发布局等具有重要意义, 对国家的科技计划制定、高新技术产业和行业认定以及企业的发展规划制定具有重要的参考价值[3]。

专利是技术创新的重要载体, 专利信息分析是突破性创新识别的重要方法之一, 主要利用专利信息中的技术知识识别突破性创新[4], 如专利分类聚类及其突变、专利主题聚类及其突变、专利权人合作以及跨界合作变化、被引专利结构和被引专利类别突变等[5]。然而, 在一些与基础研究密切相关的技术领域, 如纳米科学、基因工程、医学和生物科技等领域, 科学知识的突变或科学原理的变化为引导技术创新和突破技术瓶颈发挥了更重要的作用[6], 需要从专利信息中的被引科学知识出发对突破性创新进行识别。如第二次工业革命的多数技术发明都是建立在科学理论的基础之上的, 像电磁感应导致一系列电学应用技术出现、X射线导致X光照相技术产生等。技术新颖性是突破性

通讯作者: 张金柱, ORCID: 0000-0001-7581-1850, E-mail: zhangjinzhu@njust.edu.cn。

*本文系国家自然科学基金青年基金项目"基于被引科学知识突变的突破性创新动态识别及其形成机理研究"(项目编号: 71503125)、教育部人文社会科学研究青年基金项目"异构知识网络中主题突变动态识别研究"(项目编号: 14YJC870025)和中央高校基本科研业务专项资金项目"基于专利引用科学知识突变的突破性创新动态识别方法与形成机理研究"(项目编号: 30915013101)的研究成果之一。

创新的重要特征，而被引科学知识新颖性是技术新颖性的重要影响因素[7]，因此，可以利用科学关联度、专利引用的新科学论文数量、专利科学引文的时间分布等指标[8-9]，从专利科学引文整体上计算专利对新兴科学知识的依赖程度并以此表示技术新颖性，依赖程度越高，越可能产生突破性创新；另一方面，从被引科学知识的内容出发，通过比较不同时间段中专利科学引文的关键词差异程度识别突破性创新发生的时间以及代表性关键词[10]，还需进一步从被引科学知识的主题突变角度识别突破性创新发生的具体技术主题和技术领域。突破性创新的另一特点是多学科交叉，即平时不发生相互联系的知识领域之间发生了重组，更可能产生突破性创新[11]，一般通过专利分类号组合识别技术交叉融合[12]，而运用专利科学引文的所属学科分类组合识别突破性创新发生的技术领域还需进一步探索。

因此，本文抽取专利科学引文的关键词和学科分类及其关联关系表示被引科学知识，在关键词共现网络和学科分类组合中识别被引科学知识的主题，提出基于新关键词和重复关键词、基于新学科分类组合和重复学科分类组合的主题突变程度计算方法，并遴选突变程度高的主题对应作为突破性创新发生的技术主题，对突破性创新进行识别，完善和补充基于专利信息的突破性创新识别指标和方法。

2　利用被引科学知识主题突变识别突破性创新

如果当前专利所依据的被引科学知识主题相对于以前专利发生了突变，说明当前专利可能应用了新的科学原理、技术和方法，那么这些专利可能成为突破性创新，因此，本文以关键词和学科分类表示被引科学知识，通过对被引科学知识进行主题识别，并计算不同主题的突变程度，识别最有可能产生突破性创新的技术主题和技术领域，研究思路如图 1 所示。

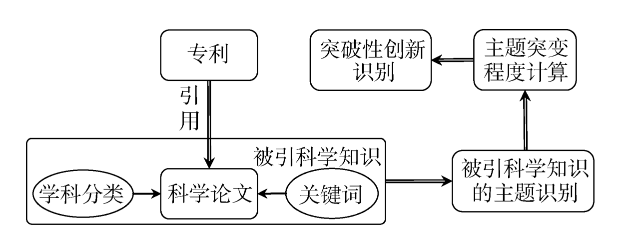


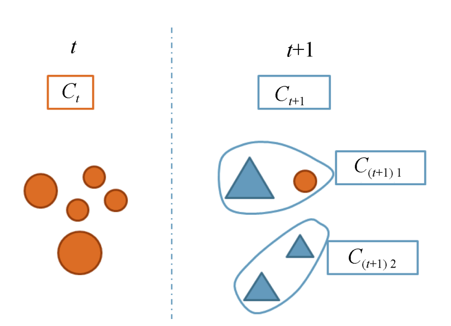
图 1　利用被引科学知识主题突变识别突破性创新的研究思路

2.1　关键词共现网络中的被引科学知识主题突变

通过对关键词共现网络进行聚类，得到某一技术领域的多个关键词主题，对这些主题的突变程度进行计算，遴选突变程度高的主题作为可能产生突破性创新的技术主题，对突破性创新进行识别。由于是专利所依据的被引科学知识突变，因此，关键词共现不是指关键词在同一专利科学引文中出现，而是指在同一专利的所有专利科学引文中出现。主题识别应用 Louvain 社团结构划分算法[13]对关键词共现网络进行聚类，下文主要对以关键词表示的主题突变程度计算方法进行说明，包含基于新关键词的主题突变程度计算和基于重复关键词的主题突变程度计算。

(1) 基于新关键词的主题突变程度计算

相对于前一时间段，当前时间段某一被引科学知识主题中包含的新关键词越多，该主题的突变程度越高，表明引用该主题的技术创新的新颖性越高，与该主题相关的技术更有可能产生突破性创新。如图 2 所示，在 t 和 t+1 时间段，C_t 和 C_{t+1} 分别表示每个时间段的所有关键词，$C_{(t+1)i}$ 表示 t+1 时间段中聚类得到的 i 主题的关键词集合，以 $C_{(t+1)i}$ 为对象说明其突变程度



$$新词数量突变率 = 1 - \frac{|C_{(t+1)i} \cap C_t|}{|C_{(t+1)i}|}$$

$$新词频次突变率 = 1 - \frac{\sum_{k \in C_{(t+1)i} \cap C_t} w_{t+1}(k)}{\sum_{k \in C_{(t+1)i}} w_{t+1}(k)}$$

$$重复词频次突变率 = \frac{\sum_{k \in C_{(t+1)i} \cap C_t} (w_{t+1}(k) - w_t(k))}{\sum_{k \in C_{(t+1)i}} w_{t+1}(k)}$$

图 2　主题突变程度计算

的计算方式。新关键词的出现频次同样会对突变程度计算产生影响，频次表示出现该关键词的专利科学引文数目，分别以 $w_t(k)$ 和 $w_{t+1}(k)$ 表示关键词 k 在 t 和 t+1 时间段的出现频次。t 时间段的关键词以圆形表示，t+1 时间段的关键词以三角形表示。

①新词数量突变率：以 t+1 时间段，某一主题中新关键词的数量占该主题中所有关键词数量的比例来表示。新关键词的数量所占比例越大，主题突变程度越高；新关键词的数量所占比例越小，主题突变程度越低。相对于 t 时间段的所有关键词 C_t，t+1 时间段 i 主题中重复出现的关键词数量为 $|C_{(t+1)i} \cap C_t|$，因此 i 主题新关键词的数量为 $|C_{(t+1)i}|-|C_{(t+1)i} \cap C_t|$，并以 i 主题的关键词总数 $|C_{(t+1)i}|$ 进行归一化，得到新词的数量突变率为 $(|C_{(t+1)i}|-|C_{(t+1)i} \cap C_t|)/|C_{(t+1)i}|$，即：

$$\text{新词数量突变率}=1-|C_{(t+1)i} \cap C_t|/|C_{(t+1)i}| \quad (1)$$

②新词频次突变率：频次突变是在数量突变的基础上，考虑主题中每个新关键词的出现频次对突变程度的影响。主题中新关键词的频次总和所占比例越大，突变程度越高；新关键词的频次总和所占比例越小，突变程度越低。相对于 t 时间段的所有关键词 C_t，t+1 时间段 i 主题中重复出现的关键词频次之和为 $\sum_{k \in C_{(t+1)i} \cap C_t} w_{t+1}(k)$，因此 i 主题新关键词的频次之和为 $\sum_{k \in C_{(t+1)i}} w_{t+1}(k) - \sum_{k \in C_{(t+1)i} \cap C_t} w_{t+1}(k)$，并以 i 主题的关键词频次总和 $\sum_{k \in C_{(t+1)i}} w_{t+1}(k)$ 进行归一化，得到新词的频次突变率为 $\left(\sum_{k \in C_{(t+1)i}} w_{t+1}(k) - \sum_{k \in C_{(t+1)i} \cap C_t} w_{t+1}(k)\right) \Big/ \sum_{k \in C_{(t+1)i}} w_{t+1}(k)$，即：

$$\text{新词频次突变率}=1-\sum_{k \in C_{(t+1)i} \cap C_t} w_{t+1}(k) \Big/ \sum_{k \in C_{(t+1)i}} w_{t+1}(k) \quad (2)$$

(2) 基于重复关键词的主题突变程度计算

相对于前一时间段，当前时间段某一被引科学知识主题中包含的重复关键词的频次变化越大，该主题的突变程度越高，表明引用该主题的技术创新可能取得突破性进展，与该主题相关的技术更有可能产生突破性创新。

重复关键词的主题突变程度通过重复词频次突变率来计算，主题中重复词的频次变化越大，突变程度越高；重复词的频次变化越小，突变程度越低。如图 2 所示，相对于 t 时间段的所有关键词 C_t，t+1 时间段 i 主题中重复词频次变化为 $\sum_{k \in C_{(t+1)i} \cap C_t} (w_{t+1}(k)-w_t(k))$，并以 t+1 时间段 i 主题的关键词频次总和 $\sum_{k \in C_{(t+1)i}} w_{t+1}(k)$ 进行归一化。此处若以 t+1 时间段 i 主题中的重复词频次总和 $\sum_{k \in C_{(t+1)i} \cap C_t} w_{t+1}(k)$ 进行归一化，则可能造成包含少量频次变化较大的重复关键词的主题突变程度过高，如 i 主题共包含 100 个关键词，仅有 1 个重复关键词并且其频次从 t 时间段的 1 增加到 t+1 时间段的 100，此时该主题的频次突变率为 (100–1)/100，这明显不符合实际情况。最终得到重复词的频次突变率为：

$$\sum_{k \in C_{(t+1)i} \cap C_t} (w_{t+1}(k)-w_t(k)) \Big/ \sum_{k \in C_{(t+1)i}} w_{t+1}(k) \quad (3)$$

2.2 学科分类组合中的被引科学知识主题突变

以被引科学知识的所属学科分类组合表示专利所依据的知识领域重组，对学科分类组合的突变程度进行计算，遴选突变程度高的学科分类组合作为可能产生突破性创新的技术领域，对突破性创新进行识别。其中，每条专利引用一条或多条专利科学引文，而每条专利科学引文具有对应的学科分类，这些学科分类便形成了学科分类组合。与关键词共现网络中的主题突变程度计算类似，学科分类组合中的主题突变同样包含两种计算方式，即基于新学科分类组合的突变程度计算和基于重复学科分类组合的突变程度计算。

(1) 基于新学科分类组合的主题突变程度计算

相对于前一时间段，当前时间段某一新学科分类组合的出现次数越多，与该学科分类组合对应的被引科学知识主题的突变程度越高，表明引用该主题的技术创新的交叉融合度越高，与该主题相关的技术领域更有可能产生突破性创新。如图 3 所示，在 t 和 t+1 时间段，C_t 和 C_{t+1} 分别表示每个时间段的所有学科分类组合，$C_{(t+1)i}$ 表示 t+1 时间段中的 i 学科分类组合，以 $C_{(t+1)i}$ 为对象说明其突变程度的计算方式。学科分类组合的出现频次表示出现该组合的专利数目，分别以 $w_t(i)$ 和 $w_{t+1}(i)$ 表示学科分类组合 i 在 t 和 t+1 时间段的出现频次。t 时间段的学科分类以圆形表示，t+1 时间段的学科分类以三角形表示。

t+1 时间段中新学科分类组合的出现次数越多，主题突变程度越高；新学科分类组合的出现次数越少，主题突变程度越低。基于新学科分类组合的主题突变程度以新学科分类组合的频次突变率计算如图 3 所示，以 t+1 时间段新学科分类组合的出现次数 $w_{t+1}(i)$表示，其中 $i \in C_{t+1} - C_t$，即：

$$\text{新学科分类组合的频次突变率} = w_{t+1}(i) \quad i \in C_{t+1} - C_t \quad (4)$$

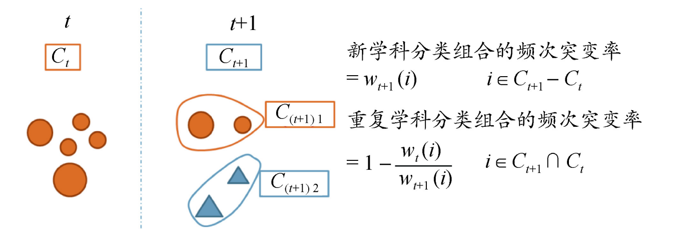


图 3 学科分类组合的突变程度计算方式

(2) 基于重复学科分类组合的主题突变程度计算

相对于前一时间段，当前时间段某一重复学科分类组合的频次变化越大，与该学科分类组合对应的被引学科知识主题的突变程度越高，表明引用该主题的技术创新可能取得突破性进展，与该主题相关的技术领域更有可能产生突破性创新。

基于重复学科分类组合的主题突变程度通过重复学科分类组合的频次突变率计算，主题中重复学科分类组合的频次变化越大，突变程度越高；重复学科分类组合的频次变化越小，突变程度越低。如图 3 所示，相对于 t 时间段的所有学科分类组合 C_t，t+1 时间段重复出现的学科分类组合 i 的频次变化为 $w_{t+1}(i)–w_t(i)$，其中 $i \in C_{t+1} \cap C_t$，并以 t+1 时间段学科分类组合 i 的出现频次 $w_{t+1}(i)$进行归一化，得到重复学科分类组合的频次突变率为 $(w_{t+1}(i) - w_t(i))/w_{t+1}(i)$，即：

$$\text{重复学科分类组合的频次突变率} = 1 - \frac{w_t(i)}{w_{t+1}(i)} \quad i \in C_{t+1} \cap C_t \quad (5)$$

3 纳米电子学领域实证分析

基于突破性创新的特征，验证领域的选择需要遵循以下三个标准：有其代表性的突破性创新，并已被权威数据证实；对科学知识的依赖程度较高，即该领域的科学关联度较高；热点的前沿研究领域，因为突破性创新会在热点的前沿领域更多的发生。

在纳米电子学领域，包括纳米导线、以碳纳米管和纳米导线为基础的逻辑电路以及只用一个分子晶体管的可计算电路组合而成的纳米电路被认定为突破性创新，在 2001 年 *Science* 杂志公布的“年度十大突破”中排名首位[14]；同时，纳米技术的专利科学引文比例超过平均水平，表明该领域发展至今其基于科学的特质，纳米技术领域对基础科学研究的依赖程度较大[15]；并且纳米技术是当前的热点研究领域，而纳米电子学是纳米技术的前沿领域[16]。基于此，本文以纳米电子学领域的数据对基于被引科学知识主题突变的突破性创新识别方法进行验证，并把纳米电路作为纳米电子学领域的突破性创新，如果该领域的被引科学知识主题突变刚好发生在纳米导线、碳纳米管、可计算电路的相关领域，则说明利用被引科学知识主题突变识别突破性创新是可能和可行的。

3.1 被引科学知识抽取

在 Thomson Innovation①中检索并下载纳米电子学领域的相关专利，筛选出包含非专利引文的专利；通过规则匹配的方式抽取非专利引文的标题；通过标题在 SCI(Science Citation Index)库中检索得到专利科学引文的关键词和学科分类。Shirabe 采用类似的方法获得专利科学引文的相关元数据[17]。

(1) 专利数据检索与获取：在 DWPI (Derwent World Patents Index)中通过专利分类号检索与纳米电子学相关的专利数据，申请日期范围为 1995 年 1 月 1 日到 2005 年 12 月 31 日，专利文献类型识别代码为 A1、A2、A9、B1、B2、E、H，检索到的专利总数为 9 359 条。相应的专利检索表达式为：

EC=((B82Y001000)) AND ADB>=(19950101) AND ADB<=(20051231) AND KI=(A1 or A2 or A9 or B1 or B2 or E or H) AND AC=(us)

(2) 非专利引文的标题抽取：剔除掉不包含非专利引文的专利，9 359 条专利中包含非专利引文的专利数目为 4 723 条，约占专利总数的 50%(4 723/9 359)，对应的非专利引文数量为 34 577 条，去除网页、专利申请以及无法识别的非专利引文后其数目为 32 601

①http://www.thomsoninnovate.org/.

条；接着识别非专利引文的标题，通过标题匹配的多种规则从 32 601 条非专利引文中成功识别出 29 355 条包含标题的非专利引文，标题的具体匹配规则和相应规则下识别出的标题数量如表 1 所示，匹配时按照类型编号的顺序进行。

表 1　非专利引文的标题抽取规则和相应数量

类型编号	规则	数量
1	标题以””开始，以””结束	16 608
2	标题以””开始，以 8221 结束	2 252
3	标题以“开始，以”结束	4 361
4	标题以 8220 开始，以 8221 结束	281
5	标题以‘开始，以’结束	801
6	标题以“、“、'、8220 中的任一个开始，以'、”、”、8221 中的任一个结束	34
7	标题以 et al.,或 et al.;或 et al.开始，以;,.中的任一个结束	2 785
8	标题以;,.中的任一个开始，以;,.中的任一个结束	2 233
9	无法识别	3 246

(3) 被引科学知识的表示：去掉标题长度为 1 的非专利引文后其数目为 29 054 条，通过非专利引文的标题到 SCI 库中进行匹配得到 15 525 条专利科学引文数据，匹配成功率为 53.4%，这些专利科学引文的关键词和学科分类及其相互关系被用来表示被引科学知识。

3.2　突破性创新发生的关键词主题识别

通过分析发现，相对于前一年，1998 年和 2001 年出现的新关键词和新学科分类组合的数量最多，二者的增长幅度也最大，导致突破性创新发生的可能性更高；同时，纳米电路在 2001 年被评为十大突破性创新之首。因此，选择 2001 年被引科学论文的关键词及其共现关系为例，对其进行聚类得到主题，并计算每个主题可能产生突破性创新的可能性，进而分析突变程度较高的主题。对 2001 年的关键词进行聚类得到 105 个主题，选择突变程度最高的主题进行说明，同时，关键词变化程度最高的前 20 个关键词均翻译成中文。

(1) 基于新关键词主题突变的突破性创新识别

依据公式(1)–公式(3)，得到第 21 个主题的新词数量突变率为 0.96，新词频次突变率为 0.92，重复词频次突变率为 0.63。该主题包含 113 个关键词，其中新词为 109 个，突变程度高的前 20 个关键词及其频次如表 2 所示。可以看到，该主题体现了化学、光学与纳米电子学的多学科交叉融合，使纳米电路的工业化应用成为可能。其中，“超分子系统”可以实现“光诱导电子转移”，从而使分子内电子转移和能量传递成为可能，为纳米电路的工业化应用提供基础条件。“富勒烯”和“二氧化硅”则是碳纳米管的重要材料和来源，碳纳米管是非常小的中空管，在电子工业有巨大的潜在应用价值，是一种新结构“富勒烯”。利用“有机化学”中的“还原氧化性”，首先在碳纳米管的内部填充金属、氧化物等物质，如“高度还原的有机金属配合物”、“有机过渡金属配合物”和“有机亲电试剂”，再把碳层腐蚀掉，就可以制备出最细的纳米尺度导线或者全新的一维材料，在未来的分子电子学器件或纳米电子学器件中得到应用。而有些碳纳米管本身还可以作为纳米尺度的导线，这样利用碳纳米管或者相关技术制备的微型导线可以置于硅芯片上，用来生产更加复杂的电路。

表 2　2001 年第 21 个主题中突变程度前 20 的关键词

关键词(频次变化)	关键词(频次变化)	关键词(频次变化)	关键词(频次变化)
还原氧化性(0,6)	Rh 纳米粒子(0,6)	有机亲电试剂(0,6)	光诱导电子转移(0,3)
高度还原的有机金属配合物(0,6)	光物理特性(0,6)	自发辐射(0,6)	电子转移反应(0,3)
有机过渡金属配合物(0,6)	有机化学(0,6)	二氧化硅(0,6)	射线晶体结构(0,3)
电荷转移复合物(0,6)	阳离子自由基(0,6)	桥接系统(0,6)	荧光反应(2,6)
富勒烯(0,6)	分子结构(0,6)	超分子系统(0,6)	能量转移(2,6)

第 5 个主题的新词数量突变率为 0.93，新词频次突变率为 0.88，重复词频次突变率为 0.18。该主题包含 190 个关键词，其中新词为 177 个，突变程度高的前 20 个关键词及其频次如表 3 所示。该主题体现了化学、生物学与纳米电子学的多学科交叉融合，使纳米尺度操作材料成为可能并在生物学上进行应用，也为纳米电路组成部分之一的分子晶体管研究提供了基础。其中，“化学力显微镜”和“原子力显微镜”是能够代表该主题的核心内容，原子力显微镜是一种纳米级高分辨率的扫描探针显微镜，优于光学衍射极限 1 000 倍，其关键是“探针”的尖端曲率半径处于纳米量级，

提供真正的"三维"表面图，是在纳米尺度操作材料及其成像和测量最重要的工具。它可以用来研究生物宏观分子，甚至活的生物组织，在"生物分子"和"结构生物学"中进行应用。同时，"化学力显微镜"则可以进行分子功能的设计，两种显微镜的"仪器标准化"和"设备校准"是其面临的关键问题，结果生成后，还需利用多种分析技术，如"化学计量学"、"多变量校正"和"偏最小二乘回归"等方法对结果进行分析。纳米电路同样需要原子力显微镜和化学力显微镜等仪器进行纳米级操作，为纳米电路相关的材料操作和分子电路研究打下基础(如分子晶体管)。

表 3　2001 年第 5 个主题中突变程度前 20 的关键词

关键词(频次变化)	关键词(频次变化)	关键词(频次变化)	关键词(频次变化)
化学力显微镜(0,12)	设备校准(0,12)	生物分子(0,12)	血糖(0,12)
原子力显微镜(0,12)	标准化(0,12)	结构生物学(0,12)	血液(0,12)
探针(2,12)	多变量校正(0,12)	3 维(0,12)	反射率(0,12)
仪器标准化(0,12)	偏最小二乘回归(0,12)	化学计量学(0,12)	水溶液(0,12)
多元仪器标准化(0,12)	仪器模型传递(0,12)	核磁共振(0,12)	近红外光谱学(0,12)

(2) 基于重复关键词主题突变的突破性创新识别

第 22 个主题的新词数量突变率为 0.77，新词频次突变率为 0.47，重复词频次突变率为 0.78。该主题包含 111 个关键词，其中新词为 86 个，突变程度高的前 20 个关键词及其频次变化如表 4 所示，可以看到，该主题体现了化学和材料科学的多学科交叉融合。在"材料科学"中，主要是以"原子力显微镜"和"扫描探针显微镜"为工具，通过"纳米光刻"等"光刻工艺"合成多种"纳米管"和纳米线，如"二氧化硅"纳米管、"单壁碳纳米管"和"金纳米"线等。金纳米技术是基于表面等离子体激元的纳米光子学，即表面等离子体激元学，一般都封装在"二氧化硅""薄膜"中，在制造纳米光子集成电路上的潜力巨大，受到了全球庞大的微电子工业的广泛关注。纳米管和纳米线的多种合成方法则与"超分子化学"密切相关，如"激光烧蚀"、溶胶凝胶和"化学气相沉积"。而"自组装单分子膜"和"硫醇单分子膜"是一种重要的制备纳米粒子的方法，它主要是利用单分子膜作为基体，在"有机物表面"上生长有机和无机纳米粒子。这些纳米材料及其制备技术均与纳米电路中的纳米管和纳米线密切相关，为其应用和发展提供了基础条件。

表 4　2001 年第 22 个主题中突变程度前 20 的关键词

关键词(频次)	关键词(频次)	关键词(频次)	关键词(频次)
材料科学(1,16)	纳米加工(2,12)	微细加工技术(6,14)	硫醇单分子膜(0,10)
激光烧蚀(5,18)	原子力显微镜(5,25)	超分子化学(1,10)	化学气相沉积(9,35)
单壁碳纳米管(2,13)	扫描探针显微镜(1,10)	卟啉阵列(1,11)	石英晶体微天平(0,13)
金纳米(12,38)	纳米光刻(1,10)	有机物表面(0,11)	模式转移(0,10)
二氧化硅(0,25)	光刻工艺(3,13)	自组装单分子膜(3,33)	链长度依赖性(0,14)

第 20 个主题的新词数量突变率为 0.69，新词频次突变率为 0.31，重复词频次突变率为 0.74。该主题包含 413 个关键词，其中新词为 287 个，突变程度高的前 20 个关键词及其频次变化如表 5 所示。该主题的主要内容为"碳纳米管"、"纳米管"、"小管"、"微管"、"量子线"、"晶体索"和"纳米线阵列"等纳米线制备材料和计算，以及与电路相关的关键词，如"电子结构"、"电子特性"、"逻辑单元阵列"、"回路"、"场发射"和"散射"等。这些纳米管相关材料和电路相关技术均与纳米电路紧密相关。

表 5　2001 年第 20 个主题中突变程度前 20 的关键词

关键词(频次)	关键词(频次)	关键词(频次)	关键词(频次)
碳纳米管(10,107)	量子线(12,68)	电子结构(5,64)	膜剂(22,127)
纳米管(0,32)	晶体索(4,52)	场发射(11,52)	单壁(1,61)
小管(11,50)	纳米线阵列(9,56)	电子特性(1,42)	散射(2,32)
微管(10,38)	线(1,36)	逻辑单元阵列(2,44)	迁移(26,65)
增长(32,199)	纳米丝(3,36)	回路(0,50)	表面(1,64)

3.3　突破性创新发生的学科分类主题识别

与突破性创新发生的关键词主题相一致，该部分同样选择 2001 年数据并对突破性创新发生的学科分类主题进行识别，通过新学科分类组合和重复学科分

类组合识别突破性创新发生的主题，计算每个学科分类组合可能产生突破性创新的可能性，分析突变程度较高的学科分类组合。

(1) 基于新学科分类组合的突破性创新识别

依据公式(4)计算新学科分类组合的突变程度，其中频次大于 3 的新学科分类组合如表 6 所示，最有可能产生突破性创新的新学科分类组合以加粗斜体字体标识，它们是化学、能源与燃料、化学工程、石油工程等学科分类的组合以及晶体学、材料科学、光学和应用物理的组合。通过频次变化可以发现，这些学科的交叉融合使得纳米电子学领域更可能产生突破性创新。与该结果一致的是，最有可能发生突破性创新的新关键词主题中，第 21 个主题的关键词体现了化学、光学与纳米电子学的多学科交叉融合，使纳米电路的工业化应用成为可能，与第 2 个学科分类组合对应；第 5 个主题体现了化学、生物学与纳米电子学的多学科交叉融合，使纳米尺度操作材料成为可能并在生物学上进行应用，也为纳米电路组成部分之一的分子晶体管研究提供了基础，与第 8 个学科分类组合对应。这也说明了新学科分类组合对于识别突破性创新所属技术领域具有一定作用，能够为优势学科技术领域布局提供参考。

表 6　代表性的新学科分类组合及其频次变化

序号	新的学科分类组合	频次变化
1	***化学、能源与燃料、化学工程、石油工程、多学科科学***	(0,6)
2	***结晶学、材料科学、表征和测试材料科学、涂层和薄膜材料科学、光学、应用物理***	(0,6)
3	多学科材料科学、凝聚态物理	(0,4)
4	电子与电气工程、仪器与设施	(0,3)
5	材料科学、多学科、应用物理、凝聚态物理、多学科物理	(0,3)
6	计算机科学与人工智能、计算机硬件与结构、计算机理论与方法、电子与电气工程、应用物理、凝聚态物理	(0,3)
7	物化学、多学科材料科学	(0,3)
8	物化学、电子与电气工程、材料科学、多学科、应用物理、物理，原子，分子和化学、凝聚态物理	(0,3)

(2) 基于重复学科分类组合的突破性创新识别

依据公式(5)计算重复学科分类组合的突变程度，其中突变程度较高的重复学科分类组合如表 7 所示，最有可能产生突破性创新的重复学科分类组合以加粗斜体字体标识。它们分别是电子电气工程和应用物理的组合、应用物理和凝聚态物理的组合以及光学和应用物理的组合，这些学科分类组合与纳米电子学领域的结合更可能产生突破性创新。与该结果一致的是，最有可能发生突破性创新的重复关键词主题中，第 22 个研究主题主要提及纳米材料及其相关制备技术，体现了化学和材料科学的多学科交叉融合，与第 4 个重复学科分类组合对应；第 20 个研究主题主要涉及纳米管制备材料以及电路，与材料科学、电子与电气工程、应用物理和凝聚态物理相关，几乎在所有的重复学科分类组合中均有部分提及，验证了基于重复学科分类组合可以识别突破性创新发生的技术领域。

表 7　代表性的重复学科分类组合及其突变程度

序号	重复的学科分类组合(频次变化)	突变程度
1	***应用物理、凝聚态物理***(1,4)	0.75
2	***光学、应用物理***(1,4)	0.75
3	***电子与电气工程、应用物理***(16,53)	0.70
4	化学、材料科学(1,3)	0.67
5	电子与电气工程、应用物理、凝聚态物理(1,3)	0.67
6	电子与电气工程、材料科学、应用物理、凝聚态物理(2,4)	0.5
7	电子与电气工程、光学、应用物理(5,9)	0.44

4　总结和展望

利用专利信息中的被引科学知识突变识别技术突变是突破性创新识别的重要方向之一，被引科学知识的主题和学科分类能够代表突破性创新所发生的技术主题和技术领域，因此，本文以专利科学引文的关键词和学科分类表示被引科学知识，提出关键词共现网络和学科分类组合中的主题突变程度计算方法，对被引科学知识中的突变主题进行识别，并以此表示最有可能产生突破性创新的技术主题和技术领域。在纳米电子学领域进行的实验发现，突变程度高的被引科学知识主题发生在纳米导线、碳纳米管、可计算电路等与纳米电路材料和制备密切相关的技术主题中；学科分类组合突变则主要集中在“化学、能源与燃料、结晶学、表征和测试材料科学、涂层和薄膜材料科学、光学、应用物理”组合和“应用物理、凝聚态物理、光学、电子与电气工程”组合中，这些组合代表的技术领域

与技术主题间具有对应关系。该结果表示利用被引科学知识的主题突变识别突破性创新是可能和可行的，验证了所提方法的有效性，并可以在其他与科学知识密切相关的技术领域进行扩展和应用。

利用被引科学知识主题突变为突破性创新识别提供了新思路，但还存在很多方面需要进一步研究，首先，以专利科学引文的关键词和学科分类表示的被引科学知识抽取准确率和召回率还需进一步提高，如穷尽非专利引文标题的所有抽取规则，并通过特定的数据库模糊检索方式进行标题匹配，从而返回更多更准确的专利科学引文元数据。同时，专利科学引文的关键词中存在大量意义宽泛的关键词，如增长、发展等，由于本文使用的是作者关键词，如果没有作者关键词则辅以机标关键词补充，因此未对这些数据进行处理，下一步还需要加强关键词的预处理工作。其次，在解析某一主题的关键词对其研究内容进行说明时，关键词间的关联关系没有清晰地展示出来，提高了解析的难度，因此，还需要对关键词间的多种语义关系进行抽取，挖掘这些主题解决了什么技术问题，并辅以可视化技术进行展示，请教相关领域专家进行解读。第三，本文所提方法能够有效识别突破性创新并有一定的预警作用，在此基础上，发现突破性创新的形成机理对其进行预测还需进一步探索。最后，仅在纳米电子学领域进行了验证，还需要在其他领域验证该方法的有效性或针对具体数据对其进行改进。

参考文献：

[1] Christensen C M, Raynor M E. The Innovator's Solution: Creating and Sustaining Successful Growth[M]. Boston: Harvard Business Press, 2003: 2-10.

[2] 张晓林. 颠覆数字图书馆的大趋势[J]. 中国图书馆学报, 2011, 37(5): 4-12. (Zhang Xiaolin. The Trends that Will Disrupt Digital Libraries [J]. Journal of Library Science in China, 2011, 37(5): 4-12.)

[3] 陈傲, 柳卸林. 突破性技术从何而来?——一个文献评述[J]. 科学学研究, 2011, 29(9): 1281-1290. (Chen Ao, Liu Xielin. How are Radical Technologies Developed? A Literature Survey [J]. Studies in Science of Science, 2011, 29(9): 1281-1290.)

[4] Dahlin K B, Behrens D M. When is an Invention Really Radical? Defining and Measuring Technological Radicalness [J]. Research Policy, 2005, 34(5): 717-737.

[5] Verhoeven D, Bakker J, Veugelers R. Identifying Ex Ante Characteristics of Radical Inventions through Patent-Based Indicators [R]. FEB Research Report-MSI, 2013.

[6] 赵志耘, 雷孝平. 我国生物科技领域技术创新与基础研究关联分析——从专利引文分析的角度[J]. 情报学报, 2012, 31(12): 1283-1289. (Zhao Zhiyun, Lei Xiaoping. Analysis of Scientific Linkage Between China's Technology Innovation and Basic Research in Biotechnology Industry Based on Patent Citation [J]. Journal of the China Society for Scientific and Technical Information, 2012, 31(12): 1283-1289.)

[7] Arts S, Veugelers R. Technology Familiarity, Recombinant Novelty, and Breakthrough Invention[J]. Industrial and Corporate Change, 2015, 24(6): 1215-1246.

[8] Sung H Y, Wang C C, Huang M H, et al. Measuring Science-Based Science Linkage and Non-Science-Based Linkage of Patents through Non-Patent References[J]. Journal of Informetrics, 2015, 9(3): 488-498.

[9] Huang M H, Yang H W, Chen D Z. Increasing Science and Technology Linkage in Fuel Cells: A Cross Citation Analysis of Papers and Patents[J]. Journal of Informetrics, 2015, 9(2): 237-249.

[10] 张金柱, 张晓林. 利用引用科学知识突变识别突破性创新[J]. 情报学报, 2014, 33(3): 259-266. (Zhang Jinzhu, Zhang Xiaolin. Radical Innovation Identification Based on Cited Scientific Knowledge Abruption [J]. Journal of the China Society for Scientific and Technical Information, 2014, 33(3): 259-266.)

[11] Kelley D J, Ali A, Zahra S A. Where do Breakthroughs Come From? Characteristics of High - Potential Inventions[J]. Journal of Product Innovation Management, 2013, 30(6): 1212-1226.

[12] Fleming L. Recombinant Uncertainty in Technological Search [J]. Management Science, 2001, 47(1): 117-132.

[13] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast Unfolding of Communities in Large Networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008. DOI: 10.1088/1742-5468/2008/10/P10008.

[14] Service R F. Breakthrough of the Year: Molecules Get Wired[J]. Science, 2001, 294(5551): 2442-2443.

[15] Meyer M, Debackere K, Glänzel W. Can Applied Science be 'Good Science'? Exploring the Relationship Between Patent Citations and Citation Impact in Nanoscience [J]. Scientometrics, 2010, 85(2): 527-539.

[16] Palmberg C, Dernis H, Miguet C. Nanotechnology: An Overview Based on Indicators and Statistics [R]. OECD, 2009.

[17] Shirabe M. Identifying SCI Covered Publications Within Non-Patent References in US Utility Patents [J]. Scientometrics, 2014, 101(2): 999-1014.

作者贡献声明:

张晓林, 张金柱: 提出研究思路, 设计研究方案, 论文最终版本修订;
张金柱: 进行实验, 采集、清洗和分析数据, 起草论文。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: zhangjinzhu@njust.edu.cn。
[1] 张金柱. ti_origin. 专利原始数据.
[2] 张金柱. snpr_title_patent. 专利科学引文的元数据.
[3] 张金柱. skcp_patent. 被引科学知识数据.

收稿日期: 2016-05-04
收修改稿日期: 2016-05-17

Radical Innovation Identification Based on Topic Mutation of Scientific Knowledge Cited in Patents

Zhang Jinzhu[1] Zhang Xiaolin[2]
[1](School of Economics and Management, Nanjing University of Science & Technology, Nanjing 210094, China)
[2](National Science Library, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: [**Objective**] Technical fields that closely related to basic research require radical innovation identification from the content of scientific knowledge cited in patents (SKCP). [**Methods**] This paper firstly extracts keywords and subject categories of scientific references in patents to represent SKCP, then identifies topics in keywords co-occurrence network and combinations of subject categories, finally proposes the method of topic mutation degree calculation based on keywords and subject categories, to identify technical topics of radical innovation. [**Results**] In the domain of Nano electronics, Nano circuit is an approved radical innovation. The related topics about this are confirmed using proposed method including Nano wire, carbon nanotubes, computing circuit, and Nano materials and the manufacturing technologies. Moreover, the corresponding combination of subject categories is materials science, chemistry, optics, biology and applied physics. [**Limitations**] The accuracy of SKCP's extraction, preprocessing and matching needs to be improved and the generality of method needs to be validated in other areas. [**Conclusions**] This method is an important improvement and supplement of radical innovation identification based on patent information and could be extended to other technical fields that are closely related to basic research.

Keywords: Radical innovation　Scientific knowledge cited in patents　Topic mutation　Mutation rate　Nano electronics